

Andrés ENRIQUE-ARIAS (ed.): *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística del corpus*. Madrid / Frankfurt: Iberoamericana / Vervuert, 2009, 416 pp.

Tal vez uno de los cambios metodológicos más importantes que se han producido en los últimos tiempos en la investigación sobre historia de la lengua sea precisamente la configuración de los *corpus* de estudio empleados en tal cometido, o, mejor, la multiplicación de tales *corpus*, desde el punto de vista tipológico —con un aumento considerable de los tipos documentales analizados— pero también, y muy especialmente, desde el punto de vista cuantitativo, con el empleo de unas cantidades de texto ingentes e impensables hasta hace pocos años; no yerra, así, el editor del volumen aquí reseñado al indicar que «hoy en día la investigación de la diacronía de las lenguas iberorromances es prácticamente inconcebible sin recurrir a los grandes *corpus* diacrónicos, bien como fuente primaria de datos, o para validar hipótesis formuladas a partir de presupuestos teóricos» (p. 14).

Sin embargo, se hace necesario señalar que el frecuente empleo de estos *corpus* de miles de documentos y millones de palabras no se ha acompañado —en general, y con las excepciones que se presentan, por ejemplo, en la nota dos de este volumen (p. 15)— de una meditación profunda y rigurosa acerca de las nuevas situaciones que se le plantean al investigador, tanto en lo que tiene que ver con la forma de extraer de ellos los mejores resultados posibles como —menos aún— con las ventajas e inconvenientes que su empleo puede suponer para los estudiosos; se echa en falta, así, una reflexión de tipo teórico que delibere acerca de *cómo* y *para qué* utilizar los nuevos *corpus* que se ponen a disposición de la comunidad científica.

Esta reflexión es, precisamente, la finalidad confesa del volumen, que pretende «contribuir a remediar esta situación mediante un panorama del trabajo que se está haciendo a día de hoy en el campo de la lingüística de *corpus* aplicada al estudio histórico de las lenguas iberorrománicas» (p. 16); son dos, por tanto, los objetivos específicos que se desprenden de la cita anterior: por un lado, una labor de catalogación, al recopilar los diversos proyectos que, en el ámbito iberorrománico, trabajan en la creación de *corpus*; por otro, y desde el punto de vista metodológico, ofrecer una serie de deliberaciones acerca de la constitución de tales *corpus* y de su explotación para la investigación diacrónica de las lenguas.

Así pues, la obra se ha organizado en seis secciones que ofrecen diferente número de trabajos —entre tres y cinco cada una de ellas— y que son las siguientes: (1) Cuestiones filológicas, (2) Diseño de *corpus* para fines específicos, (3) Cuestiones de arquitectura informática, (4) *Corpus* y variedades iberorromances, (5) La Biblia como *corpus* y (6) Aplicaciones del uso de *corpus* en la historia del español. A la vista de lo anterior, por tanto, se puede decir que son tres —a grandes rasgos— las temáticas sobre las que versa el presente volumen: por un lado, la exposición y/o descripción de los diferentes proyectos que se están desarrollando actualmente en el mundo hispánico con vistas a la configuración de un *corpus* histórico de interés lingüístico; por otro, el análisis de cuestiones teóricas que afectan a estos mismos *corpus*; por último, varios ejemplos de su explotación para el estudio de la diacronía de las lenguas (y más en concreto, del español).

Por lo que toca a la primera de las temáticas, constituye sin duda una de las más útiles para el investigador, al ofrecer una radiografía —más o menos— completa de la situación que, en relación con la creación y explotación de *corpus*, existe por el momento en el ámbito iberorrománico; se trata, además, de la que concentra el mayor número de trabajos —en concreto doce de los veintiuno que presenta el volumen— y la que justifica la presencia del término *lenguas iberorrománicas* en el título, ya que en ella no solo aparecen descripciones de *corpus* refe-

ridos al español, sino también a otras variedades románicas de origen peninsular, como el catalán (Torruella Casañas, pp. 95-115), el asturiano (Viejo Fernández *et al.*, pp. 183-94) o el gallego (Pichel & Varela Barreiro, pp. 195-215), así como el judeoespañol (Stulic-Etchevers & Rouissi, pp. 117-34; o García Moreno, pp. 217-28).

Naturalmente, los proyectos presentados en este primer conjunto de trabajos son muy diferentes entre sí, tanto en lo que tiene que ver con su estado de desarrollo —algunos ya terminados, otros en marcha y otros aún en proceso de preparación— como con los supuestos que manejan para su concepción del *corpus* —de carácter geográfico los más, aunque no todos— y las finalidades que persiguen.

Dentro de los proyectos determinados diatópicamente, cabe señalar en primer lugar el *CODEA*, que se describe en el trabajo de Sánchez-Prieto *et al.* (pp. 26-38), no solo por sus sólidas bases filológicas, sino también por el extenso ámbito del que se ocupa (todo el dominio peninsular del español, entre la Edad Media y 1700): así, el *corpus* —de 1500 documentos, accesibles de forma libre por Internet— se caracteriza por su triple presentación (facsimilar, transcripción paleográfica y presentación crítica) y tiene por objetivo «el estudio integral del documento», en el supuesto de que «sólo puede ser valorado, leído, estudiado, editado y aprovechado en investigaciones ulteriores, desde una perspectiva integradora de los diferentes puntos de vista (disciplinas) que confluyen en su estudio» (p. 27). Del mismo modo, es muy destacable también el ingente proyecto que están llevando a cabo Calderón Campos y García Godoy (pp. 229-50), y ello por dos motivos fundamentales: por un lado, por atender a una zona poco presente en las antologías de documentos lingüísticos como es el oriente andaluz²; por otro, por su inteligente decisión de editar textos de los siglos XVIII y XIX, que va a permitir llevar a cabo el necesario estudio del español de esta época, de muy escasa presencia en la bibliografía. Los primeros resultados de este proyecto se recogen ya en este estudio: datos sobre la configuración fónica o el sistema de tratamientos de la variedad dialectal de la región, que permiten vislumbrar —siquiera someramente— la importancia que el *corpus* en desarrollo tendrá para el estudio histórico de las hablas andaluzas³.

Por lo que toca al ámbito americano, destaca muy especialmente el trabajo que presenta Arias Álvarez (pp. 55-75), donde describe el *corpus* que, en unión de su grupo de investigación, está desarrollando esta estudiosa de la UNAM para el mejor conocimiento de la historia del español novohispano: así, se desgranar en estas páginas no solo los presupuestos metodológicos en que se basa el trabajo —esto es, los datos históricos y lingüísticos que justifican la organización del *corpus*— sino también los criterios de edición de los textos que se están utilizando, coincidentes en gran parte con los del *CODEA* ya mencionado. Así mismo, también se plantean unos criterios para la edición de textos del Nuevo Mundo en el estudio de Carrera de la Red y Gutiérrez Maté (pp. 39-53), quienes presentan la vertiente americana de la red internacional *CHARTA*, que persigue la creación de un estándar para la edición de documentos hispánicos de interés lingüístico: de este modo, tras analizar los criterios empleados por otros estudiosos, los

² Como los propios autores señalan, esta escasa atención a la dialectología histórica ha sido la norma en el estudio diacrónico del español peninsular: «salvo raras excepciones, los trabajos realizados hasta ahora se han hecho sin partir de un corpus regional previo, por lo que ofrecen datos valiosos pero dispersos, con los que es difícil hablar de procesos de extensión social y contextual de un determinado fenómeno» (p. 247).

³ Múltiples parecidos con éste muestra el proyecto presentado por Abad Merino (pp. 251-66), que se centra en el sureste peninsular —en concreto, la región de Murcia— y que tiene como nota más característica su clara orientación sociolingüística, para el desarrollo de la cual establece una interesante clasificación tipológica de los documentos del *corpus* (pp. 254-60).

autores aplican los que propone CHARTA a diversos documentos caribeños de los siglos XVI al XVIII, a partir de lo cual concluyen que «pese a lo diferencial de cada zona, en conjunto es más fuerte lo común en las tradiciones escriturísticas, grafemáticas y de presentación de las acciones diplomáticas en la administración española» (p. 52)⁴.

En cuanto a los *corpus* que no se caracterizan diatópicamente, cabe señalar el trabajo de Pascual & Domínguez (pp. 79-93), que describe los materiales que están sirviendo ya como base para la redacción del magno *Nuevo Diccionario Histórico del Español*, caracteriza este *corpus* y —quizá más interesante aún— explica y argumenta las decisiones que han determinado la distribución de los materiales presentes en él, tanto desde un punto de vista cronológico —con una presencia mayor de las épocas más modernas (p. 82)— como tipológico —predominio de lo no literario sobre la literatura (pp. 92-3)— o geográfico, donde se advierte la creciente importancia que progresivamente va adquiriendo América (pp. 83-4); del mismo modo, también se puede enclavar en este epígrafe el trabajo que presenta Enrique-Arias (pp. 269-83), en el que se señalan las diferentes aplicaciones y las ventajas que, para el estudio de la historia del español, puede tener el proyecto *Biblia Medieval* —en palabras del propio autor, que lo dirige, el «primer gran corpus de acceso libre en la red compuesto de textos transcritos con un criterio uniforme y que permite posibilidad de consulta de imágenes digitales de los originales» (p. 281)—, así como algunos de los problemas que su empleo conlleva, entre los que no es el menor sin duda la representatividad de la lengua empleada en él, según indica el propio autor (pp. 278-9): «la reserva más inmediata que suscita el uso de versiones bíblicas como fuente para estudiar la historia de la lengua es hasta qué punto estamos ante textos que reflejen los sistemas lingüísticos vigentes en los diferentes momentos históricos en que fueron compuestos [...]. En el peor de los casos todo ello resultaría en una lengua muy artificiosa con poco valor como fuente de datos para el estudio de la evolución del castellano general»⁵.

Por lo que toca al análisis de diferentes cuestiones teóricas que afectan a la configuración de los *corpus*, el trabajo de Mark Davies (pp. 137-66) incide en la importancia fundamental que poseen las herramientas informáticas que facilitan su empleo en la investigación: así, el autor comienza por describir los tipos de búsquedas que se pueden llevar a cabo en distintos *corpus* —en concreto, los *corpus del español* y *do português* diseñados por este mismo estudio y el *CORDE* de la Real Academia Española— para, a partir de ahí, señalar la escasa utilidad que tiene un repositorio de textos si no cuenta con los instrumentos apropiados que permiten su explotación; en sus propias palabras, «one can have a historical corpus that is composed of hundreds of millions of words of text from several different centuries, and which represent a wide range of genres. But without an adequate architecture and interface, this data is in essence 'trapped', with little if any way of getting the data out» (p. 137). En esta misma línea, Kim Schulte (pp. 167-80) analiza precisamente qué utilidades pueden tener —o cómo se pueden emplear— aquellos *corpus* que no cuentan con las herramientas señaladas más arriba, así como

⁴ Precisamente por esto sorprende una afirmación como que «de muchos documentos americanos puede decirse que, en buena medida, se escriben al margen de las tradiciones gráficas dominantes en el ámbito hispánico» (p. 51), algo que no parece casar ni con la conclusión ya señalada ni con la propia realidad que ofrece la documentación americana.

⁵ Sin llegar a tales extremos, el mismo Enrique-Arias reconoce que «hay que tener en cuenta que en un corpus de equivalentes de traducción siempre habrá cierto grado de distorsión motivada por la influencia del texto subyacente» (p. 278), lo que en principio parece contradecir su afirmación de que «es evidente [...] que los traductores tenían una competencia nativa del castellano y supeditaban la atracción por la lengua modelo a la necesidad de producir un texto entendible que siguiera las leyes gramaticales de la lengua de la época» (p. 279).

las posibles soluciones que se pueden dar a las limitaciones y problemas que esta situación plantea al investigador.

Por último, el resto de los trabajos muestran ejemplos de la explotación de los diversos *corpus* existentes actualmente para el estudio de la evolución diacrónica del español; tales trabajos se centran muy especialmente en el campo de la morfosintaxis, pero también tocan otros aspectos de la historia lingüística, como son por ejemplo el léxico (Blake & Lee, pp. 327-41) o la tipología textual (Vincis, pp. 285-303).

Entre estos trabajos se encuentra el de Alba-Salas (pp. 343-63), que lleva a cabo un estudio acerca de la evolución histórica de las colocaciones causativas como *meter miedo* entre los siglos XIII y XX, a partir del cual concluye «que en estados anteriores de la lengua *meter* competía con *poner* en estas colocaciones, que esta competición aparentemente conllevó un proceso de difusión léxica, y que el español moderno ha resuelto dicha rivalidad histórica mediante la especialización de *meter* y *poner* para contextos colocacionales diferentes» (p. 343). Por su parte, Pato (pp. 401-16) emplea varios *corpus* de grandes dimensiones —CODEA y el *Corpus del español* de M. Davies— para analizar la historia del indefinido *alguien*, y a partir de un muy fundamentado estudio y una sólida ejemplificación no solo concluye que es necesario adelantar las primeras dataciones de este elemento hasta el siglo XIII —frente a la idea mantenida hasta el momento de que su origen se encuentra en el siglo XV—, sino que también confirma su origen claramente occidental en las hablas de la Península Ibérica (p. 413)⁶.

En relación con los anteriores trabajos, Nieuwenhuijsen (pp. 365-84) también emplea un *corpus* de gran amplitud para investigar el desarrollo diacrónico de algunos pronombres en español (*connusco* y *convusco*, *vos(otros)*, *vos/os*, etc.), pero su estudio —más allá del propio análisis en sí— tiene el mérito de plantear, además, una serie de reflexiones de gran interés acerca de los problemas y las limitaciones que la utilización de estos *corpus* puede conllevar para el investigador. En esta misma línea, el trabajo de Stolova (pp. 385-400) comienza por estudiar el empleo de la auxiliación con *ser* en el español posclásico para, a partir de ahí, señalar la pervivencia de tales usos en determinadas *tradiciones discursivas* y, como consecuencia, incidir en la importancia que los *corpus* informatizados tienen a la hora de observar estos fenómenos; en sus propias palabras, «la sobrevivencia del fenómeno se observa en una serie de tradiciones discursivas que fosilizan el uso ya abandonado en la gran mayoría de la producción lingüística. Los corpus históricos en formato electrónico son imprescindibles para poder identificar dichas tradiciones con el propósito de remediar el desajuste entre las observaciones idealizadas de las gramáticas históricas y los datos reales» (p. 398).

A la luz de todo lo dicho hasta el momento no cabe, pues, sino reiterar el interés que el volumen reseñado en estas páginas puede tener para todos los estudiosos de la historia de la lengua española: en efecto, tanto por la labor de descripción de los proyectos que actualmente están en desarrollo en el ámbito hispánico, como —muy especialmente— por las múltiples reflexiones de tipo teórico que incorpora acerca de las ventajas y los inconvenientes que se descubren a la hora de emplear los distintos *corpus* ya existentes, la obra constituye, sin lugar a dudas, una herramienta de consulta de primer orden, pues aporta no solo información sobre to-

⁶ Sorprende, no obstante, otra de sus conclusiones, relacionada con la morfología de la forma en cuestión: «si bien los ejemplos más antiguos podrían tener una acentuación aguda, la documentación de la forma *alguién* [...] no ha sido muy abundante en época medieval» (p. 413); sin embargo, dada la escasa —o nula— presencia de la acentuación gráfica en la documentación medieval, parece prácticamente imposible poder discriminar cuál es la forma, *alguien* o *alguién*, que se está empleando en los textos generados durante la Edad Media.

dos aquellos temas relacionados con los corpus y su empleo, sino también —y no es poco— nuevas sugerencias y líneas de trabajo para la investigación futura.

José Luis RAMÍREZ LUENGO